Accounting for uncertainty spatial optimization

Aerts, J.C.J.H.; Goodchild, M.F.; Heuvelink, B.B.M.

published in
Transactions in GIS
2003

DOI (link to publisher)
10.1111/1467-9671.00141

document version
Publisher's PDF, also known as Version of record

Transactions in GIS, 2003, 7(2): 211–230

Research Paper

Accounting for Spatial Uncertainty in Optimization with Spatial Decision Support Systems

Jeroen C J H Aerts
Institute for Environmental Studies
Vrije Universiteit Amsterdam

Gerard B M Heuvelink
Institute for Biodiversity and Ecosystem Dynamics
Universiteit van Amsterdam

Michael F Goodchild
Department of Geography
University of California at Santa Barbara

Abstract

Spatial decision support systems (SDSS) are designed to make complex resource allocation problems more transparent and to support the design and evaluation of allocation plans. Recent developments in this field focus on the design of allocation plans using optimization techniques. In this paper we analyze how uncertainty in spatial (input) data propagates through, and affects the results of, an optimization model. The optimization model calculates the optimal location for a ski run based on a slope map, which is derived from a digital elevation model (DEM). The uncertainty propagation is a generic method following a Monte Carlo approach, whereby realizations of the spatially correlated DEM error are generated using 'sequential Gaussian simulation'. We successfully applied the methodology to a case study in the Austrian Alps, showing the influence of spatial uncertainty on the optimal location of a ski run and the associated development costs. We also discuss the feasibility of routine incorporation of uncertainty propagation methodologies in an SDSS.

1 Introduction

It has been demonstrated in practice that simple and straightforward optimization techniques linked to a spatial decision support system (SDSS) are effective for designing land use allocation alternatives (e.g. Grabaum and Burghard 1998, Cova 1999). These

Address for correspondence: Jeroen C J H Aerts, Institute for Environmental Studies, Vrije Universiteit Amsterdam, De Boelelaan 1115, 1081 HV Amsterdam, The Netherlands. E-mail: Jeroen.Aerts@ivm.vu.nl

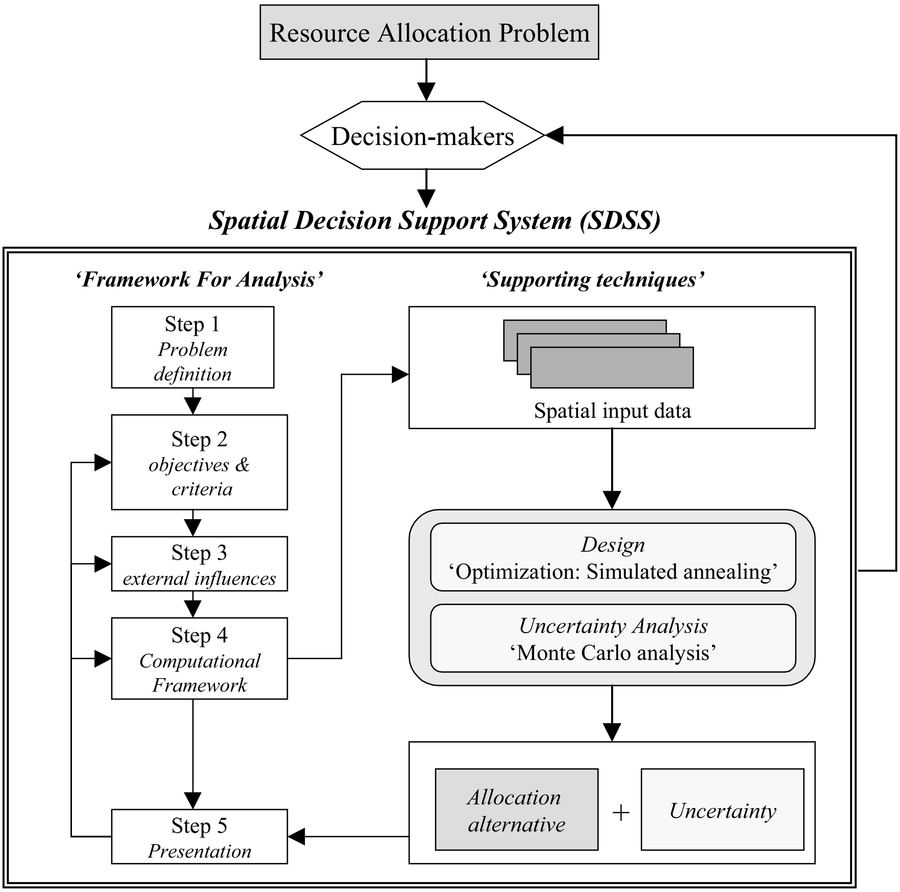


Figure 1 SDSS system with decision framework and supporting techniques

techniques can be integrated in the decision framework of an SDSS (e.g. the framework for analysis, Figure 1), and activated through a single button or slide bar. Although optimization problems sometimes are a gross simplification of 'real world' allocation problems, their results do supply a decision-maker with quick, and to a *certain extent*, reliable overviews of feasible and attractive solutions.

Above it says 'certain extent', because the solutions obtained refer to simplified problems, and they also may suffer from errors and uncertainties. Uncertainty in this context may be referred to as uncertainty in the input data, models, data interpretation and decision rules, to name a few (e.g. Brunet and Cornelis 1999, Cleaves 1995, Heuvelink 1998, Goodchild 2000, Mowrer 2000). The role of uncertainty in spatial decision-support has been pointed out by a number of studies, and although we acknowledge all forms of uncertainty as important, in this paper we restrict ourselves to the uncertainty of spatial input data. More specifically, we consider the *propagation* of uncertainty in spatial (GIS) data that affects the result of an optimization model for land use allocation.

The need for uncertainty analysis techniques within the context of spatial decision-support has been addressed by a number of studies. Hunter and Goodchild (1995) write that '. . . what is needed however is to widen the current platform of error modeling, to embrace the treatment of error from a management (or user oriented) perspective'. Furthermore, Malczewski (1999) and Agumya and Hunter (1999) state that the emergence and growth of SDSSs has promoted the use of GIS data predominantly in a conceptual way, whereas most GIS applications and subsequent uncertainty management lack the functionality to adequately support ill-structured problems. Moreover, when referring to the decision-makers themselves, Agumya (1999) states that '. . . it is recognized that although GIS data is most frequently used in operational activities, its use in higher-level decision-making is more decisive and therefore data uncertainty more harmful to the decisions being made'.

It is for this management perspective (Hunter and Goodchild 1995), for these ill-structured problems (Malczewski 1999, Agumya 1999) and for these decision levels (Agumya 1999) that we address the need for assessing uncertainty propagation within an SDSS. Any uncertainty propagation technique applied in this context should therefore meet the requirement of being *reliable*, *robust* and preferably *simple*, considering the complexity of the resource allocation issues, the often large amounts of data and great diversity in the users involved.

Monte Carlo simulation is a straightforward uncertainty analysis technique, which has been applied in a number of studies (e.g. Lee et al. 1992, Dungan et al. 1993, Journel 1996, Mowrer 1997, Fisher 1998, Heuvelink 1998, Kyriakidis et al. 1999). In this paper, we conduct an uncertainty analysis of an optimization model by applying Monte Carlo simulation using a geostatistical technique known as sequential Gaussian simulation (SGS) on the input data of the model. The objective is not to present a thorough analysis of the principles underlying geostatistics, as these are well cited in the literature (e.g. Isaaks and Srivastava 1989, Cressie 1991, Goovaerts 1997). But since the optimization model itself will be used within an SDSS for complex resource allocation issues involving non-technical users, the emphasis lies on the application. Therefore, the evaluation of the uncertainty analysis will be performed with respect to its applicability within an SDSS, by studying the following aspects:

1. *Implementation and Reliability*. This involves the implementation of Monte Carlo simulation using SGS and the exploration of the number of Monte Carlo realizations required to obtain reliable outcomes for resource allocation alternatives.
2. *Robustness*. Performance of a sensitivity analysis by changing the parameters of the variogram (sill, range) used in SGS and the analysis of how these variations affect the results of the optimization model.
3. *Significance and Simplicity*. Discussion of the practical use of uncertainty analysis within an SDSS.

These three aspects are illustrated within the context of a case study in the Austrian Alps, where the main problem is to allocate a new site for a ski run. The case study involves the use of an existing SDSS and represents a typical example of a complex resource allocation problem.

2 Methodology

2.1 Uncertainty Analysis

> 'Uncertainty in spatial data is used to denote the lack of knowledge of the true value or the value that would be discovered if one were to visit the field

and make an observation using a perfectly accurate instrument' (Hunter and Goodchild 1997).

With this definition in mind, it will be clear that almost all data stored in a GIS are uncertain to some degree. Furthermore, when the data stored in a GIS database are used as input to a GIS operation, then the uncertainties in the input will *propagate* to the output of the operation (note that from here, we use 'error' to express uncertainty in the input data). Consequently, when the propagation of the error is not adequately recorded, it becomes difficult to evaluate the accuracy of the output of the operation (Goodchild et al. 1992, Heuvelink 1998, Heuvelink 1999). For instance, a digital elevation model (DEM) will contain various kinds of error, such as measurement error, interpolation error, etc. These uncertainties will propagate when the DEM is used to compute derived products such as maps of slope, drain direction or irradiance.

2.2 Error Model

The uncertainty in a quantitative spatial attribute, such as a DEM, is typically summarized by a mean (systematic error or bias) and a variance (random error). Both can be estimated from a comparison of values of the attribute with independently collected validation data. Assuming bias is zero or corrected for, the variance, or rather its square root the standard deviation, may be estimated by the root mean squared error (RMSE). The RMSE is, however, an average value for the whole DEM and does not distinguish areas that are more or less uncertain. Furthermore, it does not assess spatial autocorrelation as present in most spatial data (Goodchild 1986). For instance, DEM values and the errors in it are most often positively correlated when measured at locations that are not too far apart (e.g. Hunter and Goodchild 1995, Fisher 1998, Goodchild 2000).

In order to perform an assessment of the propagation of error, we first define an error model of an uncertain attribute $A(x)$ at some location $x \in D$ as:

$$A(x) = b(x) + Z(x) \quad \text{for all } x \in D \tag{1}$$

where $A(x)$ is the 'true' value of the attribute, $b(x)$ is our representation of it and $Z(x)$ is the error. Due to uncertainty, the truth is unknown to some degree, and so it is represented by a random stochastic variable, which is characterized by a probability distribution. The difference between the truth and our estimate of it is given by $Z(x)$, which can be modeled as a spatially correlated random field, following:

$$Z(x) = \mu(x) + \varepsilon(x) \quad \text{for all } x \in D \tag{2}$$

Here, $\mu(x)$ is the mean of $Z(x)$ and represents the systematic error or bias, which as stated before, we take to be zero. The random field $\varepsilon(x)$ represents the non-systematic or random error. We assume $\varepsilon(x)$ to be second-order stationary and isotropic. It has zero mean and variance $\sigma^2(\mathrm{x})$, and its spatial auto-correlation is characterized by the (semi-) variance. We further assume that the semivariance is only a function of the distance $|h|$ (or 'lag') between locations (e.g. Hunter and Goodchild 1995, Heuvelink 1998, Burrough and McDonell 1998). It is defined as:

$$\gamma_z(|\mathrm{h}|) = \frac{1}{2}E[(Z(x) - Z(x + h))^2] \tag{3}$$

where E stands for mathematical expectation. A graph of the semi-variance against distance is known as the (semi-) variogram.

2.3 Uncertainty Propagation with Monte Carlo Simulation

There are a number of methods for tracing the propagation of quantitative error in spatial operations. Examples are Taylor series approximation, Rosenblueth's method and Monte Carlo simulation (Hunter and Goodchild 1995, 1997; Heuvelink 1998; Nackaerts and Govers 1999).

The Monte Carlo method is attractive for its general applicability and ease of implementation. It involves re-running an analysis many times. Each time the analysis is repeated, a (stochastic-) variable is simulated from its probability distribution and used as input for the operation. This whole process is usually repeated between 500 and 1000 times – but sometimes more and sometimes less – producing equally likely results, from here on referred to as 'realizations'. These realizations are stored and finally subject to an analysis of deriving the mean and variance across all realizations. The Monte Carlo method is often used as a method for quantifying the propagation of data base uncertainty through different operations. An operation could be a model, such as an optimization model using a DEM or slope map as input. The Monte Carlo method does not require knowledge of how the data are used in an operation and therefore this quality makes it suitable for a broad class of applications (Mowrer 1997). However, the computational load in terms of data storage and CPU time can be a major drawback (Heuvelink and Burrough 1993, Heuvelink 1998, Agumya 1999).

Monte Carlo applications for assessing the propagation of DEM uncertainty can be found by Fisher (1991a, b, 1992, 1998), Goodchild et al. (1992), Lee et al. (1992), Journel (1996), Mowrer (1997), Heuvelink (1998), Kyriakidis et al. (1999) and Holmes et al. (2000). Other examples of Monte Carlo error propagation analyses are found by Oliver et al. (1989a, b), Dungan et al. (1993), Gotway (1994), De Genst et al. (2001) and Heuvelink and Burrough (2002).

2.4 Application to DEM Uncertainty Analysis

In a spatial context, the Monte Carlo method requires the stochastic simulation of uncertain spatial attributes. Sequential Gaussian simulation (SGS) is a basic technique used for stochastic simulation in a situation where errors are normally distributed (Goovaerts 1997). SGS is, moreover, generically applicable to a vast number of applications (Journel and Deutsch 1992, Hunter and Goodchild 1995, Burrough 1999, Mowrer 2000).

The principle of Monte Carlo analysis on a spatial model using an uncertain DEM and a set of Ground Control Points (GCPs) as input, works as follows. We calculate the error at each of the GCP locations by subtracting the DEM value from the GCP value. To derive the error Z at each location x in the area of interest, we apply a sequential Gaussian simulation. Next, the spatial model is run on the DEM realizations. The procedure can be decomposed in three stages (Figure 2):

Stage A: Variogram modeling

The first Stage involves modeling the variogram of the error field $Z(x)$. First, an experimental variogram is computed from the errors observed at the GCPs. Next, we fit a function to the experimental variogram. The shape (e.g. a spherical, exponential or Gaussian) is chosen such that it optimally fits the experimental variogram.

The variogram model is controlled using the characteristic variogram parameters, i.e. the nugget, sill and range. The nugget is the variance of measurement errors combined

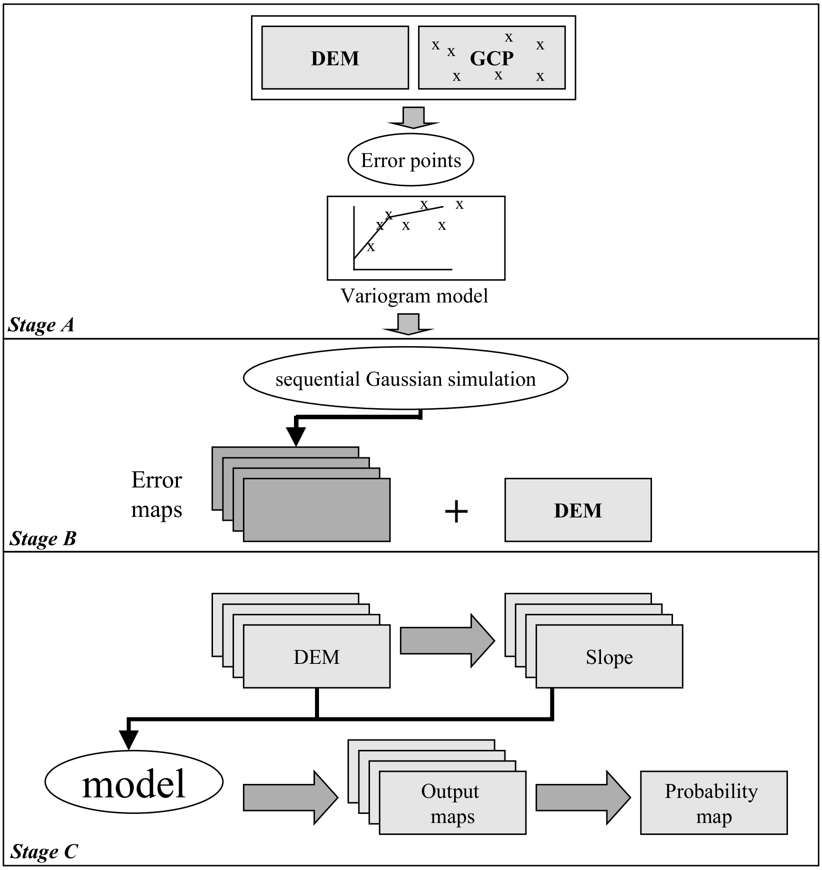


Figure 2 Monte Carlo simulation for uncertainty analysis, using sequential Gaussian simulation (SGS) in three stages

with spatial variation at distances much shorter than the sample spacing. The sill is the maximum value of the semi-variance, and equals the variance $\sigma^2(x)$. The variogram reaches the sill at a finite distance (the range), beyond which there is no longer spatial autocorrelation.

Stage B: Error map realizations

Within Stage B, SGS is applied to generate equally probable realizations of the error field $Z(x)$. For this, the SGS algorithm randomly visits each location of the area. If it is a GCP location, the value for the observed error is maintained. If not, the values of existing neighbor cells (GCP values and already simulated values) are used in a kriging interpolation to this location. As defined here, SGS may be referred to as *conditional* Gaussian simulation, since the whole procedure is 'conditioned' to the GCPs (this, as opposed to un-conditional Gaussian simulation). Kriging provides a variance and mean for $Z(x)$ (Mowrer 1997, Burrough and McDonell 1998). Then, based on that mean and variance, we assume a normal probability distribution, and a value is simulated by random selection from the normal distribution. When every location has been visited, the realization is stored and the procedure may be repeated by following a new random path through all the cells, generating a new realization. This is done N times. Finally, each error map realization is added to the original DEM, generating a set of N equally probable DEMs.

Stage C: Creation of probability maps
Within the final Stage C, the set of N DEMs is used as input for an operation, e.g. a simulation model. If the model uses a slope map as input, then first the set of DEMs is used for the calculation of N slope maps. With the set of slope maps, the model is executed N times, producing N output maps. From this, the probability distribution of the output can be derived.

3 Ski Run Planning with Optimization

We now briefly describe a generic optimization model for land use allocation using only a slope map as input.

3.1 Basic Model

The optimization model used in this paper optimally allocates new land use to an area at the lowest cost. The method divides the area in a grid, measuring N rows by M columns. Let there be K potential land uses k ($k = 1 \ldots K$). A binary variable x_{ijk} is introduced which equals 1 when land use k is assigned to cell (i, j) and equals 0 otherwise. The proportion for each new land use type is represented by parameter P_k. Thus, the sum of all cells for which $x_{ijk} = 1$ covers $P_k \times 100$ percent of the total area. Furthermore, development costs (C_{ijk}) are involved with each land use type k dependent on specific physical attributes of the area. In this paper, the only physical attribute is the slope map of the area derived from the DEM.

Because we want to minimize the development costs of the new land use, the problem may be written as an optimization model where an objective function (Equation 4 below) is minimized subject to a set of constraints. The details of the model are not further discussed in this paper (see Aerts 2002 or Aerts and Heuvelink 2002 for additional details).

$$\text{Minimize:} \quad \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{j=1}^{M} C_{ijk} x_{ijk} \tag{4}$$

This optimization model is solved using the simulated annealing algorithm, and generates *one* optimal allocation alternative per optimization run. Simulated annealing is a fast and robust optimization technique, capable of solving large combinatorial problems such as land allocation. Details on this technique can be found by Aarts and Korst (1989), Brookes (1997), Boston and Bettinger (1999), Aerts (2002) and Aerts and Heuvelink (2002).

3.2 Application of the Basic Model: Ski Run Planning

The basic optimization model has been applied to an allocation problem, which involves finding an optimal location for a ski run. The number of land uses K is 2, since each cell is either part of the ski run or not. The width of the ski run is fixed to exactly 10 adjacent cells. The model starts with allocating an initial random run between fixed start

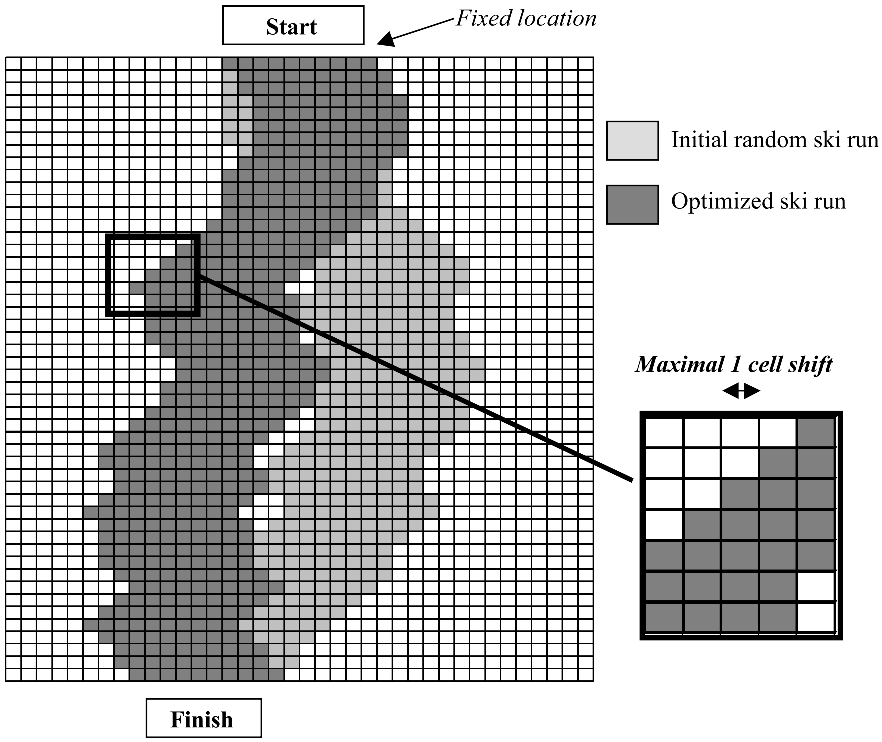


Figure 3 Grid with an initial randomly allocated ski run before optimization (*light gray*) and an optimal allocated ski run (*dark gray*) having identical start and finish points

and finish points located at opposite sides of the grid (Figure 3). This means that the total surface of the ski run would amount to 100 × 10 = 1000 cells in the case of a grid measuring 100 × 100 cells. Finally, going from the top to the bottom of the ski run, each new row of 10 ski run cells may not divert more than one cell to the left or right from that of the row above.

Development costs are the main input of the model and are assumed to be a function of the slope, as higher slopes involve more costs for artificial leveling and curving. The cost function adds both basic costs for leveling of all slope gradients (removing of rocks, smoothening) with costs for filling and curving for slopes steeper than 10 degrees. The cost function is developed by experts (Cartesian 2000) and given in Equation 5:

$$\begin{aligned} C &= a && \text{if} \quad 0 \le slope \le 10 \\ C &= a + b^{*}\, slope && \text{if} \quad slope > 10 \end{aligned} \tag{5}$$

where C (in $\$m^{-2}$) refers to the total development costs. The parameter values '*a*' and '*b*' are set to 750 and 95, respectively.

Given these conditions and cost specifications, the simulated annealing algorithm starts by creating an initial random allocation of a run between the start and finish points, and calculates the accompanying cost. Next, the model starts searching for an alternative run with lower costs. This procedure is repeated many times while minimizing development costs, until presumably the optimal ski run is obtained. Presumably refers to the fact that heuristic optimization techniques as simulated annealing do not guarantee an optimal solution (Aerts and Heuvelink 2002). However, the model was tested on its consistency by running the model many times using the same input. Each time, the model generated exactly the same result.

4 Implementation: Case Study 'Silvretta Nova'

4.1 Introduction to the Case Study

The Silvretta Nova ski resort in Vorarlberg, Austria (Plate 7) has been selected as a case study involving an authentic ski resort management problem. The study focuses on selecting a site for a new ski run at the lowest development costs. The whole planning process is a complex problem involving many stakeholders who should be consulted and informed. Because of this, the Silvretta Nova ski region decided to develop an SDSS named 'Cartesian' (Plate 7) in order to convey the new plans and create awareness of the impacts of the plans among the stakeholders in the region (Cartesian 2000). The instrument is meant to facilitate and structure the discussion on the pros and cons of ski run development in order to capture the reactions of those who will be affected by the new plans. It finally supports the planning process by selecting the most preferred site for ski run development.

The SDSS is composed of five decision steps, according to the framework for analysis (Figure 1) (Aerts 2002). Step 1 presents the problem with background information, followed by defining decision criteria (Step 2), external influences (Step 3) and finally the design of ski runs and their presentation, respectively are conducted in Steps 4 and 5.

Step 4 of the Cartesian SDDS is referred to as the 'computational step', where alternative ski run locations are generated. The simulated annealing model described in Section 3 fits into this step. We propose to implement the Monte Carlo procedure within the same computational Step 4, in order to provide the user with uncertainty information about each generated ski run. We will not elaborate on visualization of uncertainty (Step 5), but it is acknowledged as an integral part of using uncertainty analysis within an SDSS (see, for example, MacEachren 1994, Kraak 1999, Aerts 2002).

4.2 Monte Carlo Implementation

The area assigned for ski run development has been marked in the white rectangle in Figure 4. The slope map of the planning area measures 100×100 cells and is processed from a DEM using the SURFACE module in IDRISI (Eastman 1997). Costs are directly derived from the slope map, using Equation 5, and are shown with the original DEM and slope map in Figure 4. All maps have a resolution of 25×25 m^2. The start and finish locations are indicated in the same figure, and from here this proposed track is referred to as ski run Alternative 1.

Stage A

First, the error has been calculated at 70 available GCP locations in the area, by subtracting the DEM value from the independently measured GCP value. Next, these error values are used as input for modeling the variogram using the statistical software package GSTAT (Pebesma 1999). We fitted the variogram using a weighted least squares

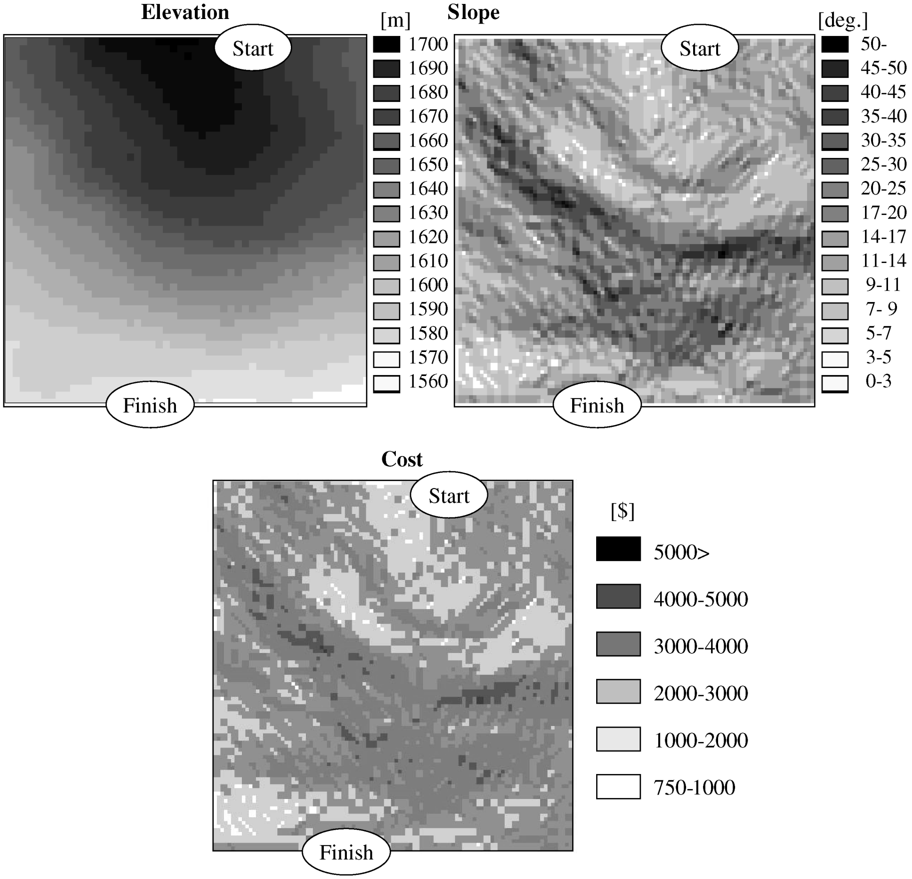


Figure 4 DEM, slope and development costs maps of the Silvretta Nova case study area with the proposed start and finish locations

minimization. The basic variogram model has a spherical shape and is shown in Figure 5. It has a range value of 2000 and a sill value of 65. Because of the smoothness of the true elevation and the DEM, we assume the nugget to be 0 as the nugget variation is very small compared to spatially dependent variation.

Stage B

For Stage B, GSTAT is used to generate a sequence of error map realizations using SGS in combination with the variogram model determined in Stage A. The whole procedure results in *N* equally probable DEMs, where *N* was initially set to 500, which is a commonly employed number in similar studies (e.g. Mowrer 1997).

Stage C

Finally within Stage C, each DEM realization is processed with the SURFACE module of Idrisi (Eastman 1997) to derive a slope map, thus yielding 500 equally probable slope maps. Next, each individual slope map becomes the input for the simulated annealing model for ski run optimization. One optimization result consists of a 0–1 realization, where a cell allocated as a ski run is assigned ‘1’, and ‘0’ otherwise.

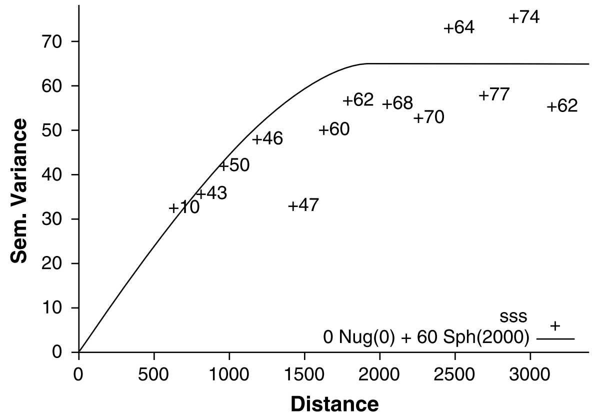


Figure 5 The basic variogram model of the error field fitted in GSTAT, using a spherical shape with a range of 2000 and a sill of 65

5 Results

5.1 Probability Results

Probability figures indicate the chance that a cell is assigned as a ski run. The probability is calculated by adding all *N* 0–1 realizations and dividing the result by *N*. But before calculating the final probability map, we first have to estimate the required value for *N* that provides a reliable estimate of the probability (remember that so far *N* has been set to 500). A reliable estimate is defined as the number of realizations at which the *variance* of the probability shows a stable value, also known as the 'point of convergence' (Mowrer 1997). For this, all 0–1 realizations are added at an increment of twenty. From this, the probability is calculated by dividing this sum by twenty for the first twenty realizations, by forty for the first forty realizations, and so on. Next, the standard deviation of the probability at each increment is calculated and plotted against the number of realizations. Figure 6 shows the results for this calculation (see 'Basic run'), where the x-axis depicts the number of realizations and the y-axis the standard deviation of the probability. From Figure 6, it appears that the standard deviation value stabilizes between 400 and 500 realizations. The initial estimate of $N = 500$ seems therefore reasonable.

Figure 7 shows the probability map as a result of the Monte Carlo analysis (from here called 'Basic run 1'). Low probabilities are expressed with dark colors and are predominantly found to the left and right (near Locations a and c). The middle section, near point 'b', is apparently an uncertain area compared to the start and finish points. Here, the model does not indicate an obvious route for the ski run. However, a closer examination of the middle section shows a slight preference for a left oriented route as depicted by lighter colors. In this area, probabilities seem to decrease going from the left to the right.

Not surprisingly, areas with the highest probability are found near the start and finish locations. These locations are fixed in the model constraints, and leave little space for the model to find many alternative routes close to those points.

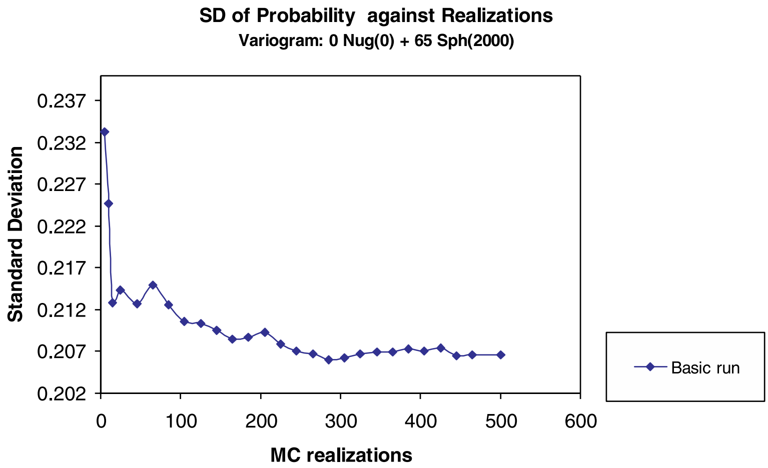


Figure 6 Average standard deviation (SD) calculated across successive increases in the number of realizations at increments of 20 realizations

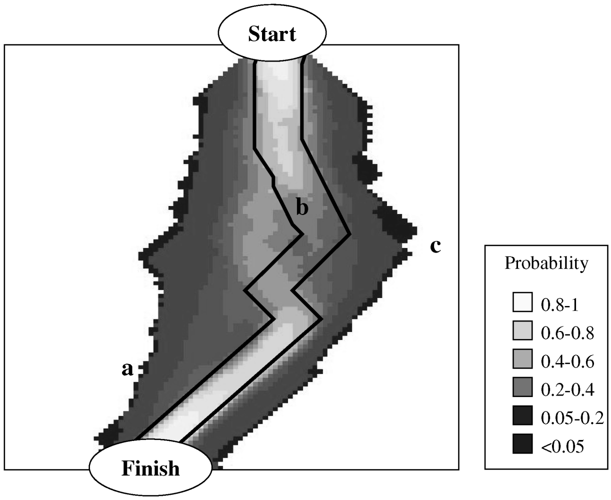


Figure 7 Probability map based on 500 Monte Carlo realizations using the basic variogram. The ski run calculated only on the basis of the original slope map, is depicted with black lines

5.2 Data Uncertainty and Development Costs

In Figure 8, the optimal track for a ski run based *only* on the original slope map is shown with black lines – thus the optimal track without accounting for uncertainty as within the Monte Carlo analysis. It appears that this track follows a different path, somewhat more to the right, especially in the middle section of the area. An interesting aspect for the user of the SDSS is to quantify the impact of these differences on the variation in development costs.

Figure 8 shows two optimal allocated ski runs for different start locations. The first (Alternative 1) is the one as discussed above. The second (Alternative 2) starts to the left

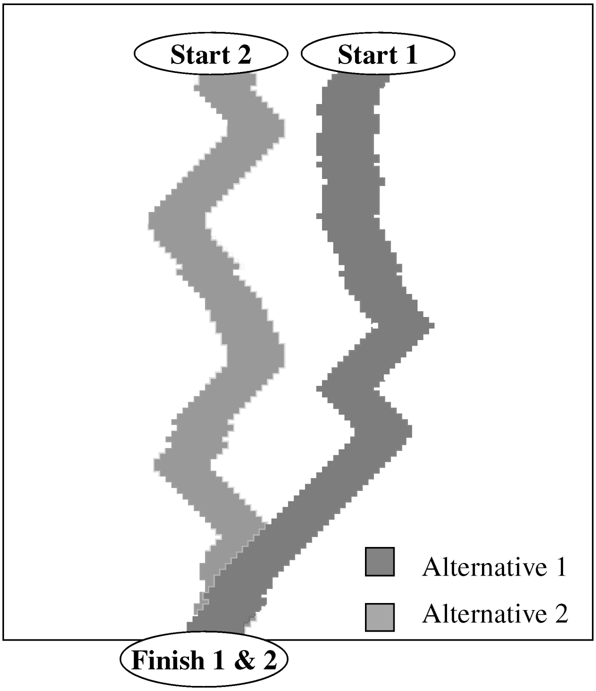


Figure 8 Alternative sites 1 (*dark gray*) and 2 (*light gray*) for a ski run. The development costs for each of these runs has been calculated across all 500 realizations and on the basis of the original slope map

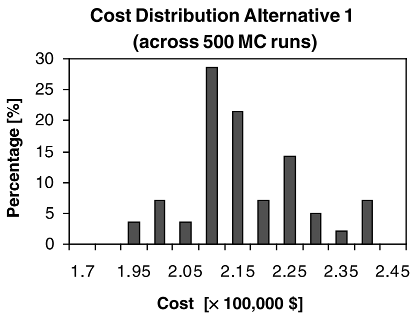
of Alternative 1. Both are required to have the same finish location. The accompanying costs for both Runs on the basis of the original slope map, are calculated to \$1,620,000 (Alternative 1) and \$1,501,000 (Alternative 2).

Next, we calculate the *expected* development costs for the same two Alternatives. Therefore, the cost distribution of the two alternatives across all 500 realizations is calculated – thus calculating the costs for the two ski run tracks for each realization *i*. The expected costs are calculated using Equation 6:

$$E = \frac{1}{500}\sum_{i=1}^{500}[f(a, b, slope_i)] \tag{6}$$

where E is the average expected costs across all 500 realizations and the cost function $f(a, b, slope_i)$ refers to Equation 5. The results are shown in Figure 9 and Table 1.

It appears that the average expected costs for Alternative 1 are \$2,135,000 and for Alternative 2 \$1,942,000. From both these numbers and Table 1, it can be concluded that the average costs calculated in the Monte Carlo analysis are higher than those calculated on the basis of the original slope map. The average difference amounts to \$441,000 and \$515,000, which indicates that most probably, the true development costs will be much higher than those based on the original slope map. This can be explained by the fact that the DEM is no longer unrealistically smooth, but is transformed through the Monte Carlo Analysis, in a more realistic DEM with higher variability. Consequently, this has resulted in a slope map with higher values, yielding higher



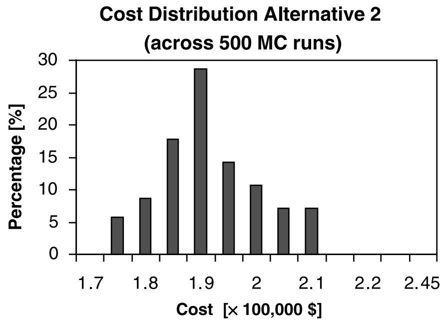


Figure 9 Histogram of the development costs across all 500 realizations

Table 1 Development cost [$]

	Original slope map	Average MC	Difference
Alternative 1	1,620,000	2,135,000	515,000
Alternative 2	1,501,000	1,942,000	441,000

costs using Equation 5. Hence, an investment in a more detailed and accurate DEM (and thus a more accurate slope map), could enhance the accuracy for finding an optimal route for a ski run and its related cost.

5.3 Joint Probability

In previous sections, it was concluded that the left oriented route for Alternative 1 seems to be the most probable route. The probability map supports an *indication* for this observation by depicting slightly lighter colors towards the left of point b (Figure 7). However, the probability numbers are individual values per cell, and do not indicate whether cells are jointly allocated as a ski run within one 0–1 realization.

The joint probability (JP) approach calculates the probability for two cells being *jointly* allocated as ski run cells across all 0–1 realizations. For this, the uncertain middle section was subdivided into three potential ski run routes: *left*, *middle* and *right* (Figure 10). The JP can be estimated for a cell pair (U, V) placed in each of those three routes with the expression:

$$P\{U = u \ \& \ V = v\} = \frac{\sum_{k=1}^{N} R_k}{N} \qquad (10)$$

Here, U and V are individual cells that lie in the uncertain region. N is the number of realizations and R_k refers to the number of realizations that meets the condition ($U = u$ & $V = v$). R_k equals 1 if $U = u$ and $V = v$, and 0 otherwise. Since we are looking for cells being allocated as a ski run within each 0–1 realization, we set u and v to 1.

The results of the JP calculation for the Basic run 1 show indeed the highest JP value of 0.55 for the cell pair to the left. This implies a preference for a left oriented route.

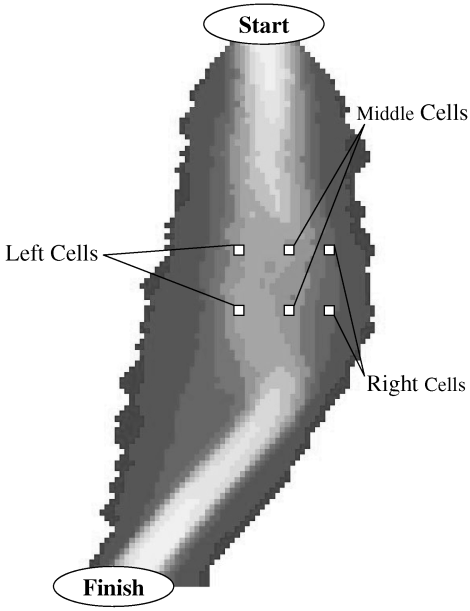


Figure 10 The joint probability has been calculated for three pairs of cells that occur in the routes of three possible runs: *left, middle* and *right*

The cell pair at the right shows the lowest JP value of 0.29, and hence is the least probable route for a ski run. The result for the cell pair in the middle lies in between those of the left and right pairs, but shows a JP value closer to the result for the left cell pair. The latter is again an indication that a left oriented route is preferred when accounting for the uncertainty of the input data.

5.4 Robustness Analysis: Consequences for the Probability

With the variogram used for the Basic run 1 (Figure 5), we now successively increase and decrease the values of the range and sill parameters. Each changed parameter value yields a new variogram model, which in turn is used for a new Monte Carlo analysis. The range has been assigned to a value of 2500 (Run 2) and 1500 (Run 3) keeping the sill at a value of 65. Thereafter, the range value has been kept constant at 2000 while varying the value of the sill at 50 (Run 4) and 100 (Run 5).

Figure 11 shows the standard deviation values of the probability for the different Monte Carlo Runs 2 to 5 as opposed to the Basic run 1. Runs 2 to 5 show a constant value for the standard deviation of the probability between 400 and 500 realizations. Furthermore, it can be derived that compared to the basic run (sill = 65, range = 2000), both a higher range and a lower sill result in a lower variation in the probability. The opposite is shown as well, as a lower value for the range and a higher value for the sill result in higher values for the standard deviation.

Table 2 shows the variation in parameter values and, consequently, the percentage change after 500 realizations of the standard deviation (SD) as compared to the Basic run. An interesting aspect is that although differences are small, an *increase* in the range

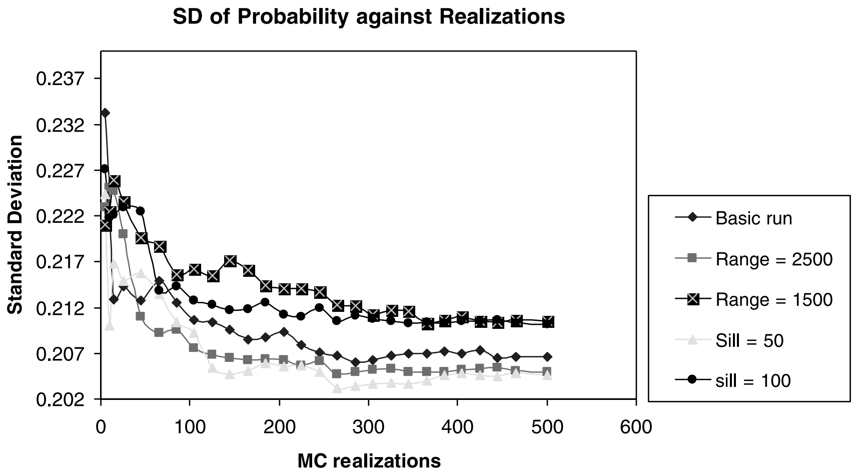


Figure 11 Average standard deviation (SD) calculated across successive increases in the number of realizations at increments of 20 realizations

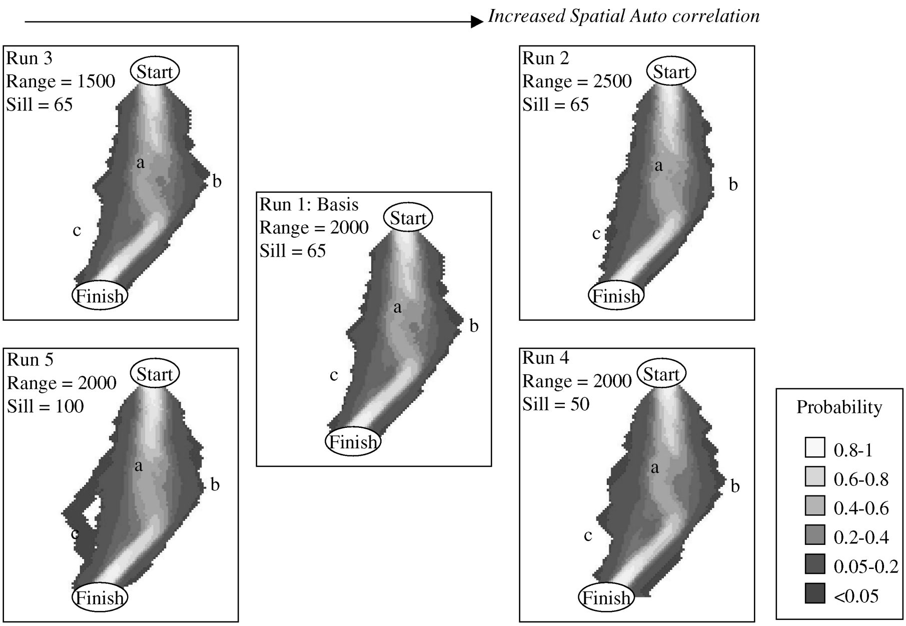


Figure 12 Probability maps for all Runs. The spatial auto-correlation increases from left to right. Higher probabilities are indicated with lighter colors

value of 25% (2000 to 2500) lowers the SD by 0.9%, whereas a *decrease* in the range by 25% (2000 to 1500) increases the SD by 1.9%. This is more or less twice as much. Variations in sill values show that a decrease of 23% in the sill value (65 to 50) lowers the SD by 0.9%, whereas an increase in the sill value by 100% (50 to 100) results in an increase of the SD by 1.7%.

Table 2 Parameters used for variogram modeling in ski run case

Run	Variable	Model	Nugget	Range	Sill	SD	*% change*
run 1: Basic	Probability	Spherical	0	2000	65	0.207	=
run 2	Probability	Spherical	0	2500	65	0.205	−0.8
run 3	Probability	Spherical	0	1500	65	0.211	1.9
run 4	Probability	Spherical	0	2000	50	0.205	−0.9
run 5	Probability	Spherical	0	2000	100	0.210	1.7

5.5 Spatial Consequences of Parameter Variations

We now examine the spatial consequences of the variation in variogram parameters. Figure 12 depicts the probability per cell (across all 500 0–1 realizations) for being included as a location for a ski run, for Runs 1 to 5. In general, it can be seen that the probability field narrows with increased spatial autocorrelation (higher range values), thus going from the left (Run 3), via the Basic run 1, to the right (Run 2).

In order to make a closer examination of spatial differences, we use Locations a, b and c in Figure 12. Cells close to the fixed start and finish points (e.g. down to Location c) show obviously the highest probability as the simulated annealing model does not have much latitude to allocate ski run cells at a distance from these fixed points. Again, for all Runs, the left-oriented route (thus left from Location a) seems to be more preferable for being allocated as a ski run according to the somewhat lighter colors.

The difference in the probability field width due to variation in the range is clearly shown at Location b. When comparing Runs 2 and 3, it can be derived that Run 3 nearly 'touches' Location b as opposed to Run 2. A similar observation can be made by comparing the variation in probabilities due to different sill values. When comparing Runs 4 and 5 at Location c, Run 5 clearly shows outliers of allocated ski runs to the left of Location c indicating an increased variation in the model results and thus a relatively high variation in the input data.

6 Discussion and Conclusions: Practical Use Within an SDSS

Uncertainty analysis for SDSSs is still a relatively unexplored area. This is partly due to the technical character of uncertainty analysis methods, which are therefore misunderstood or considered too difficult by decision-makers using an SDSS. Furthermore, uncertainty *can* play many roles in the decision-making process, with different kinds of impacts. That is often used as an excuse for not dealing with it. A better view would be to examine all of its possible impacts, just as we need to examine all possible sources of uncertainty in input data. To date, there have been few examples of the application of uncertainty methods to problems that exhibit a decision-making character, while a few studies point out the importance of doing exactly that (e.g. Hunter and Goodchild 1995, Cleaves 1995, Malczewski 1999, Agumya 1999).

Agumya (1999) describes the *significance* of uncertainty analysis for higher-level-decision-making. We therefore used a case study in Austria to demonstrate the use of uncertainty analysis within an SDSS, developed for stakeholders involved in the ski run planning process, including representatives from the government, tourist business, forestry institutes and environmental agencies. They are here referred to as higher-level-decision-makers, using the SDSS in a workshop environment.

Agumya (1999) describes uncertainty as risks, and explicitly states that uncertainty, such as data error, may have a greater impact for higher-level-decision-making compared to lower-level-decision-making (e.g. technical field experts). This can be illustrated for the Silvretta Nova case study where a workshop was organized focusing on either including or excluding potential ski run sites based on broadly defined criteria such as development costs and environmental impacts. The uncertainty information described in this paper is one of the aspects that may have influenced this process. Once it has been decided which ski run alternative is preferred, technical experts will determine where to exactly construct the new run, but within the boundaries given by the decision-makers at the workshop.

We would like to point out that some studies warn that uncertainty should not be defined as risks (Cleaves 1995). Many decision-makers that were present in the above-described workshop are not likely to accept (data-) uncertainty as a risk because it could undermine the publics' support for the new plans. Moreover, not very uncommon in this respect is the use of similar GIS data by different stakeholders in the decision process, claiming different policies to be undertaken due to uncertainty in the problems underlying GIS information (Hunter 1999). This stresses the fact that uncertainty information is not only more *decisive* at higher decision levels, it also may add political *sensitivity* to the whole planning process when it is not carefully managed.

If we picture a user of an SDSS as described above, then most likely, this user does not want to be confronted with detailed information about variogram parameters, and other statistical details. However, the merit of the uncertainty method is simple and straightforward. In this respect, it is assumed that each user understands the mechanism by which a perturbation in input data will have an effect on the outcome of a calculation using those data. The described Monte Carlo approach is therefore considered as applicable to an SDSS but we suggest detailed information to be excluded, while the different probability maps that show the user the boundaries of the feasible planning area, are clearly presented.

The probability information and its derivative impact on other criteria, such as cost, is the most valuable information for a user of an SDSS. This information can be especially effective for the elimination of alternative sites. Elimination is one of the most commonly employed approaches for decision-makers in the area of resource allocation in order to reduce a huge set of possible sites into a surveyable number of sites.

As pointed out by other research, it can be argued that the total calculation time of the whole uncertainty approach is still a drawback for a direct implementation within an SDSS. The total calculation time for one Monte Carlo run amounts to 4–5 hours on a Pentium III-450 MHz computer. However, the bulk of the calculation time is accounted for by re-running the simulated annealing model. The whole Monte Carlo analysis, including the GSTAT calculations and the calculation of probability maps, is a matter of minutes. It is expected that within several years the whole methodology might run in a shorter and reasonable amount of time in an SDSS.

Acknowledgements

The data used were processed and provided by the Cartesian consortium under the project EU-DGXII/CEO, project no *ENV4-CT-98-46*. We thank Diederik Waardenburg for implementing the simulated annealing algorithm. We furthermore thank Ashton Shortridge (Michigan State University), Phaedon Kyriakidis (University of California at Santa Barbara) and two anonymous referees for valuable comments.

References

Aarts E and Korst J 1989 *Simulated Annealing and Boltzman Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. New York, John Wiley and Sons

Aerts J C J H 2002 Spatial Decision Support for Resource Allocation: Integration of Optimization, Uncertainty Analysis and Visualization Techniques. Unpublished PhD Dissertation, University of Amsterdam

Aerts J C J H and Heuvelink G B M 2002 Using simulated annealing for resource allocation. *International Journal of Geographic Information Science* 16: 571–87

Agumya A 1999 Risk-Based Approach for Assessing the Fitness for Use of Geographical Information. Unpublished PhD Dissertation, Department of Geomatics, University of Melbourne

Agumya A and Hunter G 1999 Translating uncertainty in geographical data into risk in decisions. In Shi W, Goodchild M F and Fisher P F (eds) *Proceedings of the International Symposium on Spatial Data Quality*, Hong Kong: 574–84

Boston K and Bettinger P 1999 An analysis of Monte Carlo integer programming, simulated annealing, and tabu search heuristics for solving spatial harvest scheduling problems. *Forest Science* 45: 292–301

Brookes C J 1997 A parameterized region-growing program for site allocation on raster suitability maps. *International Journal of Geographical Information Science* 11: 375–96

Brunet S and Cornelis B 1999 Uncertainties in spatial decisions: A policy maker point of view. In Shi W, Goodchild M F and Fisher P F (eds) *Proceedings of the International Symposium on Spatial Data Quality*, Hong Kong: 566–73

Burrough P A 1999 GIS and Geostatistics: Essential partners for spatial analysis. In Shi W, Goodchild M F and Fisher P F (eds) *Proceedings of the International Symposium on Spatial Data Quality*, Hong Kong: 10–20

Burrough P A and McDonell R A 1998 *Principles of Geographical Information Systems*. New York, Oxford University Press

Cartesian 2000 CARTESIAN: A Management Information System for Ski Resorts. In Aerts J and Wilms R (eds) *Demonstration of Earth Observation Techniques*. Amsterdam, Institute for Environmental Studies, Vrije Universiteit Amsterdam, Final Research Report for Project EU-DGXII, Project No ENV4-CT-98-46

Cleaves D A 1995 Assessing and communicating uncertainty in decision support systems: Lessons from an ecosystem policy analysis. *AI Applications* 9: 87–102

Cova T 1999 A General Framework For Optimal Site Search. Unpublished PhD Dissertation, Department of Geography, University of California at Santa Barbara

Cressie N A C 1991 *Statistics for Spatial Data*. New York, John Wiley and Sons

De Genst W, Canters F, and Gulinck H 2001 Uncertainty modelling in buffer operations applied to connectivity analysis. *Transactions in GIS* 5: 305–26

Dungan J L, Peterson D L, and Curran P J 1993 Alternative approaches for mapping vegetation quantities using ground and image data. In Michener W K, Brunt J W, and Stafford S G (eds) *Environmental Information Management and Analysis: Ecosystem to Global Scales*. London, Taylor and Francis: 237–62

Eastman J R 1997 *IDRISI for Windows (Version 2.0) Users Guide*. Worcester, MA, Clark University

Fisher P F 1991a Modelling soil map-unit inclusions by Monte Carlo simulation. *International Journal of Geographical Information Systems* 5: 193–208

Fisher P F 1991b Algorithm and implementation uncertainty in viewshed analysis. *International Journal of Geographical Information Systems* 7: 331–47

Fisher P F 1992 First experiments in viewshed uncertainty: Simulating fuzzy viewsheds. *Photogrammetric Engineering and Remote Sensing* 58: 345–52

Fisher P F 1998 Improved modelling of elevation error with geostatistics. *GeoInformatica* 2: 215–33

Goodchild M F 1986 *Spatial Autocorrelation*. Norwich, Geobooks Catmog No 47

Goodchild M F 2000 Introduction: Special issue on 'Uncertainty in Geographic Information Systems'. *Fuzzy Sets and Systems* 113: 3–5

Goodchild M F, Guoqing S, and Shiren Y 1992 Development and test of an error model for categorical data. *International Journal of Geographic Information Systems* 6: 87–104

Gotway C A 1994 The use of conditional simulation in nuclear-waste-site performance assessment. *Technometrics* 36: 129–41

Goovaerts P 1997 *Geostatistics for Natural Resources Evaluation*. New York, Oxford University Press

Grabaum R and Burghard C M 1998 Multi-criteria optimization of landscapes using GIS-based functional assessments. *Landscape and Urban Planning* 43: 21–34

Heuvelink G B M 1998 *Error Propagation in Environmental Modelling with GIS*. London, Taylor and Francis

Heuvelink G B M 1999 Propagation of error in spatial modeling with GIS. In Longley P A, Goodchild M F, Maguire D J, and Rhind D W (eds) *Geographical Information Systems: Principles, Techniques, Applications and Management*. New York, John Wiley and Sons: 188–92

Heuvelink G B M and Burrough P A 1993 Error propagation in cartographic modelling using Boolean logic and continuous classification. *International Journal of Geographic Information Systems* 7: 231–46

Heuvelink G B M and Burrough P A 2002 Developments in statistical approaches to spatial uncertainty and its propagation. *International Journal of Geographic Information Science* 16: 111–3

Holmes K W, Chadwick O A, and Kyriakidis P C 2000 Error in a USGS 30m Digital Elevation Model and its impact on terrain modelling, *Journal of Hydrology* 233: 154–73

Hunter G J 1999 Managing uncertainty in GIS. In Longley P A, Goodchild M F, Maguire D J, and Rhind D W (eds) *Geographical Information Systems: Principles, Techniques, Applications and Management*. New York, John Wiley and Sons: 633–41

Hunter G J and Goodchild M F 1995 Dealing with error in spatial databases: A simple case study. *Photogrammetric Engineering and Remote Sensing* 61: 529–37

Hunter G J and Goodchild M F 1997 Modelling the uncertainty of slope and aspect estimates derived from spatial databases. *Geographical Analysis* 29: 35–49

Isaaks E H and Srivastava R 1989 *An Introduction to Applied Geostatistics*. New York, Oxford University Press

Journel A G 1996 Modelling uncertainty and spatial dependence: Stochastic imaging. *International Journal of Geographical Information Systems* 10: 517–22

Journel A G and Deutsch C V 1992 *GSLIB, Geostatistical Software Library and User's Guide*. New York, Oxford University Press

Kraak M J 1999 Visualising spatial distributions. In Longley P A, Goodchild M F, Maguire D J, and Rhind D W (eds) *Geographical Information Systems: Principles, Techniques, Applications and Management*. New York, John Wiley and Sons: 157–73

Kyriakidis P C, Shortridge A M, and Goodchild M F 1999 Geostatistics for conflation and accuracy assessment of Digital Elevation Models. *International Journal of Geographical Information Science* 13: 677–707

Lee J, Snyder P K, and Fischer P F 1992 Modelling the effect of data errors on feature extraction from Digital Elevation Models. *Photogrammetric Engineering and Remote Sensing* 58: 1461–7

MacEachren A M 1994 *Some Truth with Maps: A Primer on Symbolization and Design*. Washington, Association of American Geographers

Malczewski J 1999 *GIS and Multi-criteria Decision Analysis*. New York, John Wiley and Sons

Mowrer H T 1997 Propagating uncertainty through spatial estimation processes for old-growth subalpine forests using sequential Gaussian simulation in GIS. *Ecological Modelling* 98: 73–86

Mowrer H T 2000 Uncertainty in natural resource decision support systems: Sources, interpretation and importance. *Computers and Electronics in Agriculture* 27: 139–54

Nackaerts K and Govers G 1999 Accuracy assessment of probabilistic visibilities. *International Journal of Geographic Information Science* 13: 709–21

Oliver M A, Webster R, and Gerrard J 1989a Geostatistics in physical geography: Part I, Theory. *Transactions of the Institute of British Geographers* 14: 259–69

Oliver M A, Webster R, and Gerrard J 1989b Geostatistics in physical geography: Part II, Applications. *Transactions of the Institute of British Geographers* 14: 270–86

Pebesma E J 1999 Gstat User's Manual. WWW document, http://www.geog.uu.nl/gstat/